Quantitative Methods – Written Report

Why do employees in Berlin earn less than German average? A quantitative approach to explaining earnings differentials

Introduction

The literature on labor productivity[1] and earnings in cities is broad. Glaeser et al. (1992) review part of it and suggest that income growth is driven by knowledge spillovers[2], which are particularly strong in the dense living and working environments of cities. Bettencourt et al. (2007) confirm that knowledge spillovers spur growth. In addition, they find that "larger cities are associated with higher levels of productivity" and that wages and income "scale superlinearly with city size". Glaeser and Maré (1994) show that wage differences across space are a compensation for differences in the cost of non-traded goods (mainly land), and that these costs are again a function of local human capital. It appears that the literature can be compiled to knowledge spillovers in cities which elevate the growth of human capital[3]. This again drives productivity and hence earnings for workers in cities. The larger and denser the urban labor market the more knowledge spills over.

Berlin is Germany's capital, its largest city and most densely populated federal state (EURES, 2016). With an "economically active population" of 1,786,400 people in 2015 it is also the country's largest urban labor market (Eurostat, 2016). However, unlike theory predicts, the city's labor market has been underperforming German average for years in terms of wages, employment, and disposable incomes (Brenke, 2016; Eurostat, 2014).
There are recent studies and statistics that address wages and unemployment in Berlin (DIW, 2016; Brenke, 2016). This work aims to add to the discussion by placing its focus on employed people in Berlin, analyzing their individual labor earnings and human capital levels. It questions why employees in Berlin earn less than German average contrarily to what theory predicts.
Part one introduces the available data for analysis, part two discusses the methods employed, part four presents and discusses results and part five concludes.
The aim is to provide further insights and grounds for effective labor market policies in Berlin.

1 Data

The data employed for analysis comes from the German Socioeconomic Panel (SOEP). The SOEP is a "wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research (DIW Berlin)" (DIW, 2017). The panel study was

[1] Labor productivity is defined as the value added per employed person (Eurostat, 2013).
[2] Knowledge spillovers stand for the exchange and flow of ideas among individuals (Carolino, 2001).
[3] Human capital is defined as "the stock of skills that the labor force possesses", which is usually acquired through education and training (Goldin, 2016).

first launched in 1984 and comprises data on household composition, occupational biographies, employment, earnings, health and satisfaction indicators (DIW, 2017). The data available for this work is SOEP v31, which features the panel of 2014.
The data has been collected by TNS Infratest Sozialforschung under the name "Living in Germany", mainly through face-to-face interviews which are based on a set of pre-tested questionnaires (Glemser et al., 2015; DIW, 2017). Households are selected by random-walk (DIW, 2017).
Researchers can gain access to SOEP data through the SOEP Research Data Center. For that they have to sign a data protection contract according to German data protection laws, which we have done. After that the data was made available to us in STATA format.

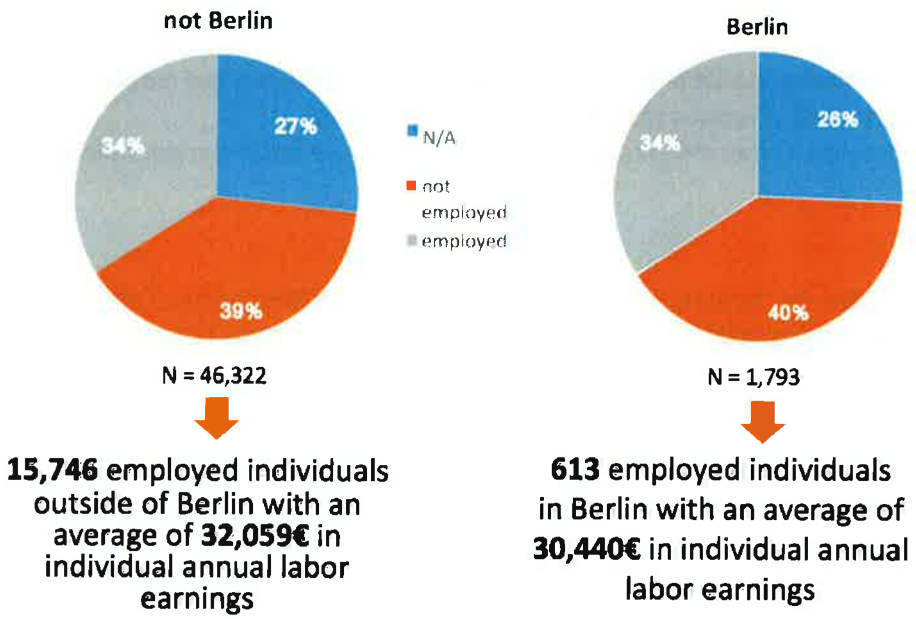
The variables that apply to the analysis of this work are state of residence, employment status, individual labor earnings, age of the individual, years of education, and gender.

Figure 1 – Selected variables of the SOEP 2014 data

Individual	State of Residence	Employment status	Individual annual labor earnings	Years of education	Age	Gender
1	Bavaria	Employed	56,250	18	58	Male
...	...	...	...	...		
48,115	Berlin	Employed	30,666	11.5	42	Female

Our sample features information on 48,115 individuals, of whom 1,793 are Berlin residents. As this work focuses on why employees in Berlin earn less than German average, only the sample of employed people is used for analysis. This leaves us with information on 16,359 individuals, of whom 613 reside in Berlin.

Figure 2 – Sample for analysis



2 Methodology

The relationship between earnings and human capital has been subject to quantitative analysis since Mincer (1958) (Solomon, 2007). Mincer proposed regression of earnings on human capital to explain why differences in the distribution of earnings among individuals exist. In labor economics human capital usually comprises measures of education and experience (Mincer, 1974). This constitutes the basic regression formula, which is known as the human capital earnings function:

Figure 3 - Conceptual human capital earnings function

$$Ln\ earnings_i = \alpha + \beta_1 * educ_i + \beta_2 * exp_i + \varepsilon$$

where α is an individual's initial earnings capacity, β_1 the percentage gain in earnings for an additional year of education, and β_2 the percentage gain in earnings for an additional year of experience[4] (Solomon, 2007). The explanatory variable educ is measured in years of education, exp is measured in age - years of education - 6 [5].

The regression model is a log-linear one because the distribution of earnings appears to be positively skewed with an increasing variance of residuals (Figures 4a and 5b). Log transformation makes the earnings distribution more similar to a normal distribution and the variability in the residuals more constant (Figures 4b and 5b). A log-transformation of our earnings data thus enables us to meet the prerequisites for linear regression (Princeton University, 2008; StatTrek, 2017; Pennsylvania State University, 2017a).

[4] In a log-linear regression model a one-unit change in the independent variable leads to a $(\beta*100)$ % change in the dependent variable (everything else held constant). For the mathematical derivation see Benoit (2011) and Cornell Statistical Consultancy Unit (2012).

[5] Children in Germany usually start school at the age of 6, so an individual's potential labor market experience is age – years of education – 6 (Mincer, 1972).

Figure 4a – Earnings distribution before log transformation

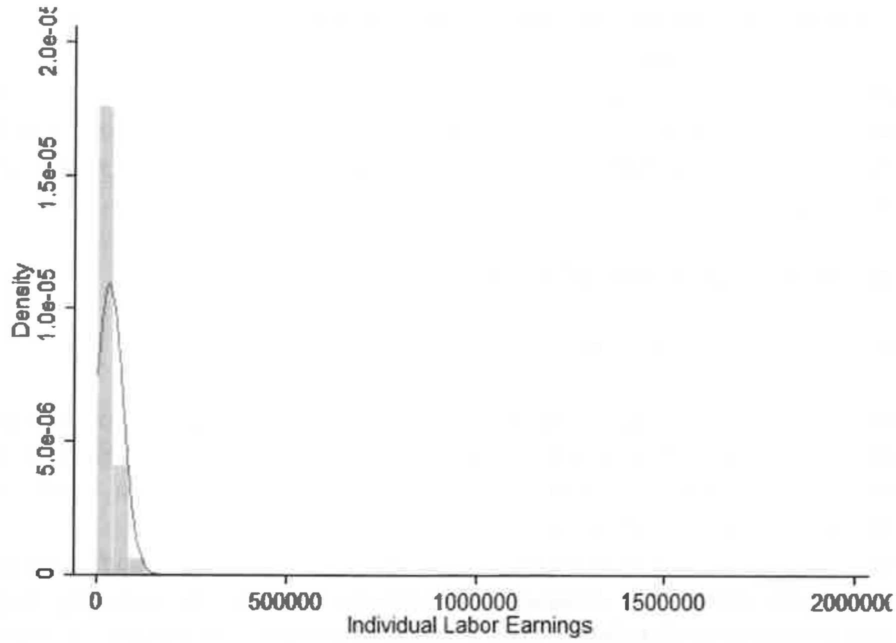


Figure 4b – Earnings distribution after log transformation

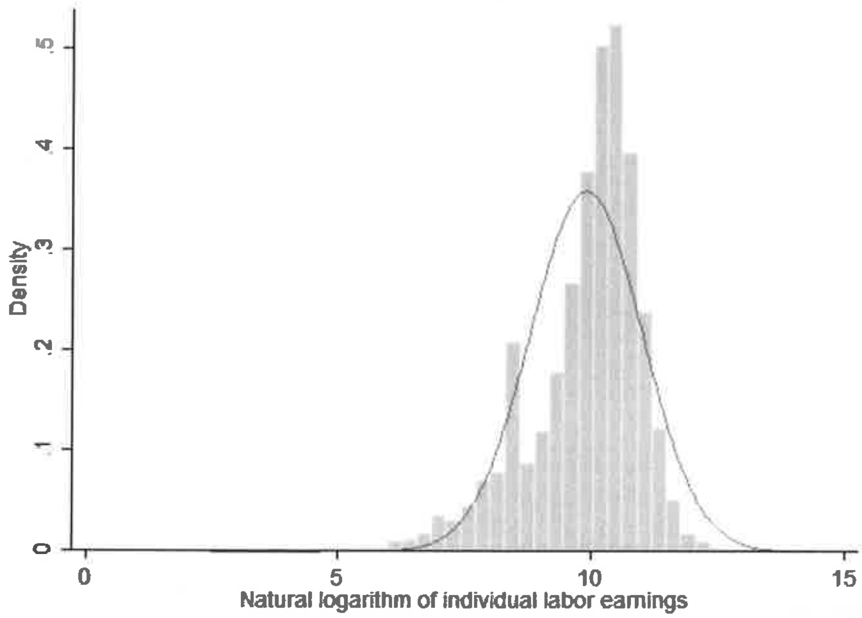


Figure 5a – Residual plot of earnings before log transformation

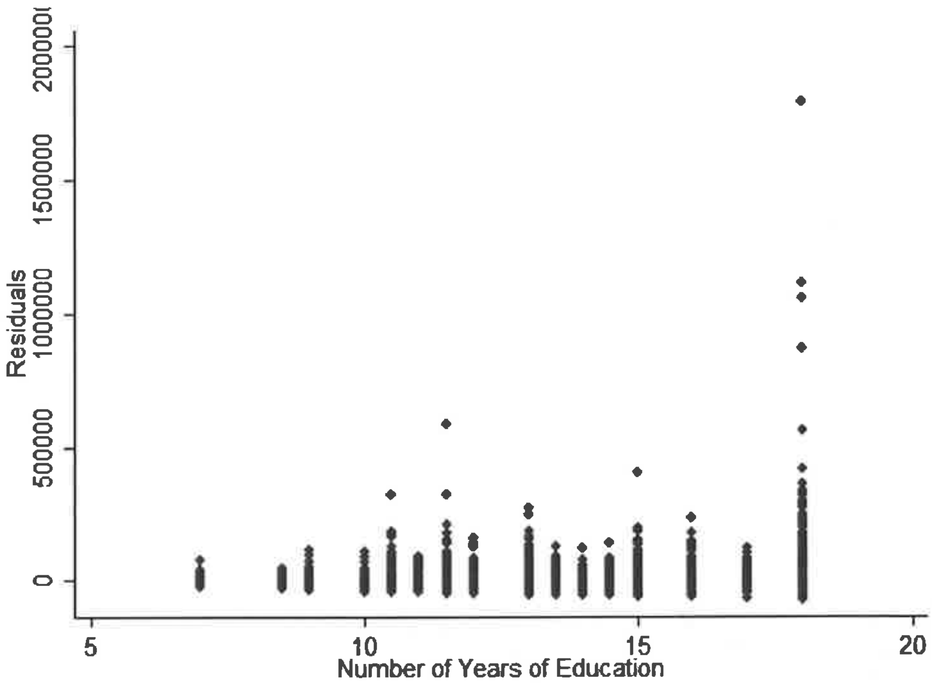
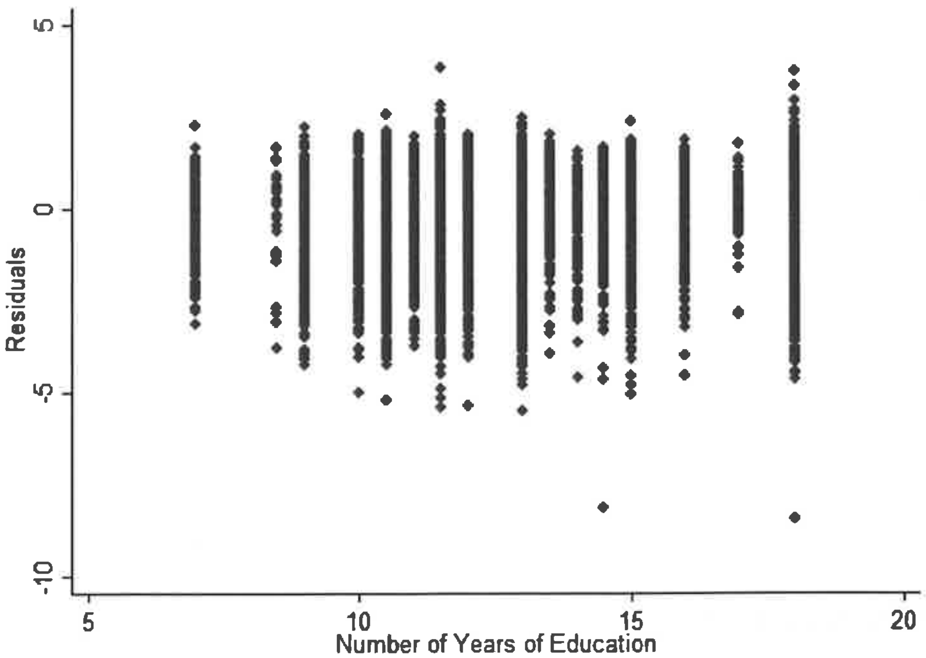


Figure 5b – Residual plot of ln(earnings) after log transformation



The proposed relationship between earnings and human capital implies two hypotheses about why employees in Berlin earn less than German average for our analysis:

1) Levels of education and/or experience are lower in Berlin.
2) Education and/or experience gets less rewarded in Berlin, expressed in lower β -coefficients for Berlin residents.

2.1 Hypothesis One

We can test hypothesis one with two one-tailed mean comparison tests. As we compare two different samples with "nearly equal" variances [6], we employ a pooled variance two-sample t-test (Pennsylvania State University, 2017b).

Figure 6 – Summary statistics human capital

	Berlin	**Not Berlin**
Mean years of education	13.97 years	12.8 years
Variance education	8.43	7.4
Std. Dev. education	2.9	2.72
Mean experience	24.86 years	25.88 years
Variance experience	155.19	132.79
Std. Dev. experience	12.46	11.5

Figure 7 – Pooled standard deviation and test statistic for the two-sample t-test

Pooled standard deviation

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Test statistic

$$t^* = \frac{mean_1 - mean_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

n_1 and n_2 – sample sizes from population 1 and 2
s_1 and s_2 – standard deviations of population 1 and 2

(Pennsylvania State University, 2017b).

[6] A rule of thumb to determine whether the variances are equal is to check whether the ratio of the two sample standard deviations falls from 0.5 to 2 which holds true for our data (Pennsylvania State University, 2017b).

2.1.1 Test of difference in means of years of education

H_o: $years\ of\ education_{Berlin} =\ years\ of\ education_{no\ Berlin}$
H_1: $years\ of\ education_{Berlin} <\ years\ of\ education_{no\ Berlin}$ $\quad ,\alpha = 0.05$

2.1.2 Test of difference in means of experience

H_o: $experience_{Berlin} =\ experience_{no\ Berlin}$
H_1: $experience_{Berlin} <\ experience_{no\ Berlin}$ $\quad ,\alpha = 0.05$

2.2 Hypothesis Two

We can test hypothesis two with a regression of individual labor earnings on years of education and experience. For that we have to make some specifications to the conceptual formula (Figure 3).
As Mincer (1972) argues in his later work the relationship between experience and earnings is not a linear but a concave one due to diminishing returns to human capital with age.

Figure 8a – Proves that relationship is not solely linear

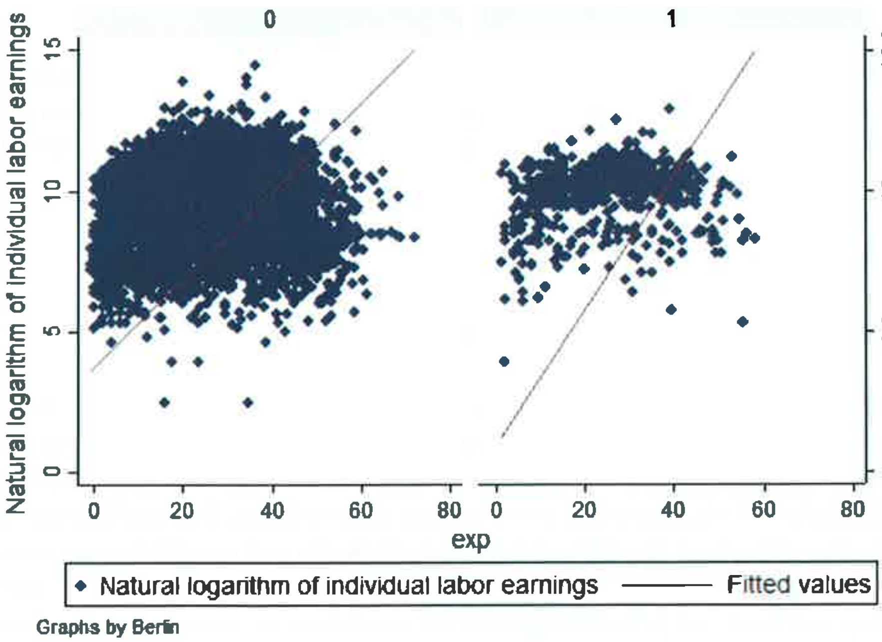
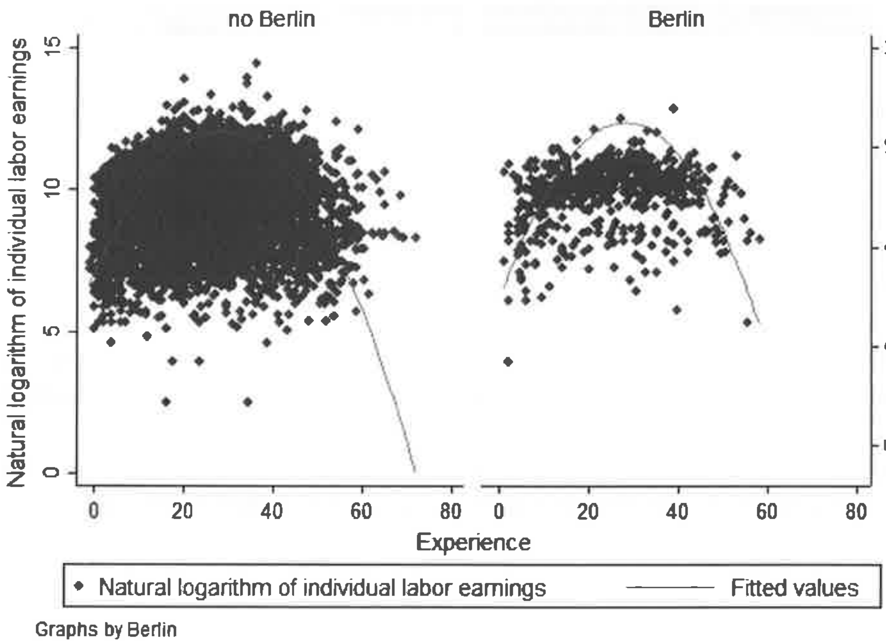


Figure 8b – Proves that a sole quadratic fit is not sufficient either



To capture the effect of experience on earnings correctly we thus have to add a squared version of experience to the human capital earnings function[7]. The experience and experience squared terms together can describe a monotonic relationship with one inflection point (ESSN, 2013a).

Figure 9 – Adjusted regression formula

$$Ln\ earnings_i = \alpha + \beta_1 * educ_i + \beta_2 * exp_i + \beta_3 * exp_i^2 + \varepsilon$$

As we are interested in how the β coefficients differ for Berlin residents, we have to make another adjustment to the regression formula and introduce interaction terms. For that we will multiply our human capital explanatory variables, namely $educ$, exp, and exp^2, with a binary Berlin variable. Berlin = 1 indicates that an individual lives in Berlin. The coefficients of the interaction terms then represent the differences between Berlin and non-Berlin residents in the effect of an additional year of education or an additional year of experience on earnings (ESSN, 2013b). Last but not least, we introduce gender as an additional control variable. This gives us the final regression model for testing hypothesis two (Figure 10).

Figure 10 – Final regression formula

$$\begin{aligned} Ln\ earnings_i = {} & \alpha + \beta_1 * educ_i + \beta_2 * exp_i + \beta_3 * exp_i^2 + \beta_5 * educ_i Berlin_i \\ & + \beta_6 * exp_i Berlin_i + \beta_7 * exp_i^2 Berlin_i + \beta_8 * gender_i + \varepsilon \end{aligned}$$

---

[7] A one unit change in experience then leads to a ($\frac{\partial(\ln(earnings))}{\partial exp} = \beta_2 + 2\,\beta_3\, exp$)*100% change in earnings.

3 Results

3.1 On Hypothesis One – Lower human capital levels in Berlin

3.1.1 H_1: $years\ of\ education_{Berlin} < years\ of\ education_{no\ Berlin}$

Difference in means	-1.175635
Std. Error	0.1141664
t	-10.2976
Pr(T>t)	1.0

The results of mean-comparison-test number one, namely diff = -1.18 and Pr(T>t) > 0.05, show that we cannot accept our hypothesis that the level of education for Berlin residents is significantly lower than it is for non-Berlin residents. In fact, the negative difference in means suggests that Berlin residents are on average higher educated.

3.1.2 H_1: $experience_{Berlin} < experience_{no\ Berlin}$

Difference in means	1.015784
Std. Error	0.483871
t	2.0993
Pr(T>t)	0.0179

The results of mean-comparison-test number two show that we can reject our null hypothesis of equal means in experience. Non-Berlin individuals are on average more experienced than Berlin residents.

3.1.3 Overall result hypothesis one

Our overall results are ambiguous. We cannot fully confirm that levels of human capital are lower in Berlin. We can, however, say that employees in Berlin are less experienced than their non-Berlin counterparts.

3.2 Hypothesis Two – Human capital gets less rewarded in Berlin

	Coefficient	Std. Error	t	P >|t|
educ	0.1314425	.0028043	46.87	0.000
exp	0.1028199	.0024436	42.08	0.000
exp^2	- 0.0017382	.0000447	-38.89	0.000
educBerlin	- 0.0212969	.0079493	-2.68	0.007
expBerlin	0.0216534	.0100811	2.15	0.032
exp^2Berlin	- 0.0004417	.0002023	-2.18	0.029
gender	- 0.7711514	.0148469	-51.94	0.000
α	7.363585	.0496559	148.29	0.000

R^2	Adj. R^2	F	Prob > F
0.2994	0.2991	963.74	0.0000

The regression results show that our model can explain 30% of the variability in earnings among individuals (R^2). The p-values indicate that all of the explanatory variables in our model are significant. Berlin residents get on average and ceteris paribus 2.13% less rewarded than non-Berlin residents for an additional year of education as expressed by the educBerlin coefficient. An additional year of experience yields an increase of 2.01% in earnings for Berlin residents compared to a 1.48% increase for non-Berlin residents[8].

As for hypothesis one, we cannot fully confirm that Berlin residents get less rewarded for their human capital. We can confirm that they get less rewarded for their education. However, additional experience leads to more gains in earnings in Berlin than it does outside of Berlin.

3.3 Limitations to the model

The gender coefficient with -0.7712 for women appears to be very high. This in addition to the relatively small R^2 of 30% suggests that there are omitted variables in our regression, whose effect was probably taken up by the gender variable. Omitted variables could, for example, be job types in different sectors or industries. As there is still a difference in job types between men and women, the high value of the gender coefficient probably incorporates part of the effect of different job types on earnings. Our model could be improved if there was additional data available on job types of individuals which we could introduce as additional controls.

4 Conclusion

Our analysis shows that education levels in Berlin are higher than outside of Berlin but that they get less rewarded there in terms of earnings. On the other hand, experience levels in Berlin are lower than outside of Berlin but get more rewarded in terms of earnings. While we cannot fully confirm that Berlin employees earn less than German average because of overall lower human capital, we can say that a difference in experience levels plays a role in explaining earnings differentials. We can assume that an increase in experience levels would benefit earnings in Berlin. This goes along with policy advice from the German Institute for Economic Research (DIW, 2016). They suggest that Berlin's start-up scene has much potential for good

[8] See footnote 7 for calculation.

labor market conditions but that the city also needs more mature companies which drive employment and productivity. It can be assumed that more mature companies would be a more attractive workplace for highly experienced workers. Policy should thus aim to support Berlin start-ups to grow into more mature companies. Our model could be improved and policy advice specified by adding controls for different job types.

References

Benoit, K., 2011. Linear Regression Models with Logarithmic Transformations. London School of Economics Methodology Institute, online available at: http://www.kenbenoit.net/courses/ME104/logmodels2.pdf, retrieved on Jan 6, 2017.

Bettencourt, L., Lobo, J., Helbing, D., Kühnert, C., and West, G., 2007. Growth, innovation, scaling, and the pace of life in cities. National Academy of Science USA, PNAS, Vol 104, No. 17, pp. 7301-7306, online available at: www.pnas.org/cgi/doi/10.1073/pnas.0610172104, retrieved on Dec 3, 2016.

Brenke, K., 2016. The Berlin labor market since 2005: strong employment growth yet unemployment remains high, incomes low. DIW Economic Bulletin 29+30.2016, pp. 327-337, online available at: https://www.diw.de/documents/publikationen/73/diw_01.c.540059.de/diw_econ_bull_2016-29-3.pdf, retrieved on Dec 9, 2016.

Carolino, G., 2001. Knowledge Spillovers: Cities' Role in the New Economy. Philadelphia Fed Business Review, Q4 2001, pp. 17-26, online available at: https://www.philadelphiafed.org/research-and-data/publications/business-review/2001/q4/brq401gc.pdf, retrieved on Jan 5, 2017.

Cornell Statistical Consulting Unit, 2012. Interpreting Coefficients in Regression with Log-Transformed Variables. StatNews #83, June 2012, online available at: https://www.cscu.cornell.edu/news/statnews/stnews83.pdf, retrieved on Dec 8, 2016.

European Social Survey Net (ESSN), 2013a. Adding non-linearity to OLS regression models. (online) available at: http://essedunet.nsd.uib.no/cms/topics/multilevel/ch1/5.html, retrieved on Dec 3, 2016.

European Social Survey Net (ESSN), 2013b. Adding interaction terms to OLS regression models. Online available at: http://essedunet.nsd.uib.no/cms/topics/multilevel/ch1/6.html, retrieved on Dec 3, 2016.

Eurostat, 2013. Glossary: Labour productivity. Online available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Labour_productivity, retrieved on Dec 7, 2016.

Eurostat, 2014. GDP at regional level. Online available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/GDP_at_regional_level. Retrieved on Dec 3, 2016.

EURES (The European job mobility portal), 2016. Labour market information, Germany, Berlin. Online available at: https://ec.europa.eu/eures/main.jsp?countryId=DE&acro=lmi&showRegion=true&lang=en&mode=text&regionId=DE0&nuts2Code=%20&nuts3Code=null&catId=375, retrieved on January 3, 2017.

Eurostat, 2016. Cities (Urban Audit), Labour market – cities and greater cities (urb_clma), last update 21-12-2016. Online available at: http://ec.europa.eu/eurostat/web/cities/data/database, retrieved on January 6, 2017.

German Institute for Economic Research (DIW), 2016. The Berlin economy and labor market: can the city transform from startup hotbed to burgeoning metropolis?. Press Release of 25 July 2016. Online available at: https://www.diw.de/sixcms/detail.php?id=diw_01.c.539669.en , retrieved on Dec 7, 2016.

German Institute for Economic Research (DIW), 2017. SOEP v31. Online available at: https://www.diw.de/en/diw_01.c.519355.en/soep_v31.html, retrieved on Jan 2, 2017.

Glemser, A., Huber, S., Bohlender, A., 2015. SOEP 2014 – TNS Report of SOEP Fieldwork in 2014. SOEP Survey Papers, Series B, 282.

Goldin, C., 2016. Human Capital. Handbook of Cliometrics, Springer Verlag, online available at: http://scholar.harvard.edu/goldin/publications/human-capital, retrieved on Jan 8, 2017.

Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A., 1992. Growth in Cities. Journal of Political Economy, vol. 100, no. 6, online available at: http://scholar.harvard.edu/files/shleifer/files/growthincities.pdf, retrieved on Jan 2, 2017.

Glaeser, E. and Maré, D., 1994. Cities and skills. NBER Working Paper No. 4728, online available at: http://www.nber.org/papers/w4728, retrieved on Nov 30, 2016.

Mincer, J., 1958. Investment in Human Capital and the Personal Income Distribution. Journal of Political Economy, 66(4), pp. 281-302, online available at: http://www.jstor.org/stable/pdf/1827422.pdf, retrieved on Dec 18, 2016.

Mincer, J., 1974. Schooling, Experience, and Earnings. Columbia University Press for the National Bureau of Economic Research, online available at: http://www.nber.org/chapters/c1767, retrieved on Dec 18, 2016.

Pennsylvania State University, 2017a. 1.3 The Simple Linear Regression Model. STAT 501 Regression Methods, Department of Statistics Online Programs, online available at: https://onlinecourses.science.psu.edu/stat501/node/253, retrieved on Jan 2, 2017.

Pennsylvania State University, 2017b. 8.3 Comparing Two Population Means: Independent Samples. STAT 501 Regression Methods, Department of Statistics Online Programs, online available at: https://onlinecourses.science.psu.edu/stat500/node/50, retrieved on Jan 3, 2017.

Princeton University, 2008. Log transformations. The Princeton University Data and Statistical Services, online available at: http://dss.princeton.edu/online_help/stats_packages/stata/log.html, retrieved on Jan 2, 2017.

Solomon, P., 2007. Earnings Over the Lifecycle: The Mincer Earnings Function and Its Applications. IZA Discussion Paper No. 3181, online available at: http://ftp.iza.org/dp3181.pdf, retrieved on Dec 9, 2016.

StatTrek, 2017. Residual Analysis in Regression. Online available at: http://stattrek.com/regression/residual-analysis.aspx, retrieved on Jan 2, 2017.